Dedicated Workspaces: Faster Resumption Times and Reduced Cognitive Load in Sequential Multitasking

Steven Jeuris[a,*], Jakob E. Bardram[b]

[a] *The Pervasive Interaction Technology Laboratory, IT University of Copenhagen Rued Langgaards Vej 7, 2300 Copenhagen, Denmark*
[b] *Department of Applied Mathematics and Computer Science, Technical University of Denmark Richard Petersens Plads, Building 324, 2800 Kgs. Lyngby, Denmark*

Abstract

Studies show that virtual desktops have become a widespread approach to window management within desktop environments. However, despite their success, there is no experimental evidence of their effect on multitasking. In this paper, we present an experimental study incorporating 16 participants in which a traditional Windows 7 environment is compared to one augmented by virtual desktops. Within the experimental condition, each virtual desktop acts as a dedicated workspace devoted to an independent goal-oriented task, as opposed to the control condition where only one single workspace is available to perform the same tasks. Results show that adopting virtual desktops as dedicated workspaces allows for faster task resumption (10s faster on average) and reduced cognitive load during sequential multitasking. Within our experiment the majority of users already benefited from using dedicated workspaces after three switches to a previously suspended task, as the time lost on setting up workspaces was compensated for by faster subsequent task resumption. These results provide a strong argument for supporting goal-oriented dedicated workspaces within desktop environments.

Keywords: window management, multitasking, dedicated workspaces, virtual desktops

1. Introduction

Virtual desktops are a widespread approach to window management within desktop environments. Since their introduction as part of the Rooms system (Henderson & Card, 1986), similar systems have been researched and deployed in commercial operating systems, including several Linux distributions and OS X (named 'Spaces'). The latest release of Windows 10 now also includes support for virtual desktops, demonstrating the relevance of gaining a better understanding on how they are used and how they influence knowledge work. Knowledge work is diverse, but includes extensive processing of information as observed in

*I am corresponding author
Email addresses: sjeu@itu.dk (Steven Jeuris), jakba@dtu.dk (Jakob E. Bardram)

e.g., design, management, law, finance, and research (Kidd, 1994). This sometimes leads to knowledge workers experiencing information overload (Mulder et al., 2006), which is where virtual desktops are posited to help out.

Despite the widespread availability and adoption of virtual desktops in desktop environments, there is currently no empirical evidence on how they influence the work of knowledge workers in terms of task performance and cognitive load. However, a few observational studies have identified emerging usage patterns when using multiple desktop environments. For example, Ringel (2003) identified five different strategies for dividing work across virtual desktops; by task, subtask, primary/secondary work, operating system, and application. Studies investigating the use of multiple monitors and multiple computers identified similar organizational strategies (Beale & Edmondson, 2007; Grudin, 2001; Hutchings et al., 2004). Organizing work in larger thematically connected units of work has also been observed within an office environment, where in addition to electronic tools the use of physical artifacts was taken into account, such as books and telephones (González & Mark, 2004). Hence, evidence shows that there is a clear need to support the intertwining of task-centric units of work (*sequential multitasking*), and users have been shown to adopt virtual desktops to this end. Therefore, many 'activity-centric' (or 'activity-based') personal computing systems use virtual desktops to support the user in grouping together application windows pertaining to the same task (Bardram et al., 2006; Houben et al., 2013; Jeuris et al., 2014; Voida et al., 2008).

In this paper we study the effects of organizing work in task-oriented *dedicated workspaces*, as opposed to a 'side-by-side' configuration where all open resources are visible and readily accessible. We conducted an experimental study measuring task resumption time, task construction time, self-reported cognitive load, task productivity, and accuracy under both conditions. Our results advocate supporting goal-oriented dedicated workspaces during sequential multitasking, as they reduce the time to switch between tasks, and the users' perceived task load considerably. We observe less experienced users benefit most from using dedicated workspaces. No significant difference in task productivity or accuracy was observed, which we attribute to a trade-off with respect to ecological validity in our methodology. We will address the implications of this, and resulting guidelines for future work, as part of our discussion.

2. Background and Related Work

The goal of this paper is to measure the effects of using dedicated workspaces during multitasking. As we will discuss here, several studies in both cognitive sciences and human-computer interaction study multitasking, as well as related concepts such as interruptions and task switching. However, some ambiguity exists in the terminology used. Where some theories, models, and studies investigate short-term subconscious processes, others describe activities which can last up to several hours. Research related to multitasking can be divided into (at least) four different topics: studies investigating (*i*) concurrent multitasking, (*ii*) task switching, (*iii*) interruptions, and (*iv*) sequential multitasking.

Although unifying theories of the full spectrum—ranging from concurrent to sequential multitasking—have been proposed (Salvucci et al., 2009), the individual empirical studies do not report on the same phenomena and are hence hard to interpret as a whole (Janssen et al., 2015). Our study focuses on investigating sequential multitasking, but to situate it within this wider research, and to prevent misinterpreting study results, we provide a short overview and report on important differences and similarities.

2.1. Concurrent Multitasking

During concurrent multitasking cognitive resources have to be divided across several competing parallel tasks, such as driving while talking on the phone. Many 'dual-task' studies investigate dual-task interference and reduction of performance while performing two simultaneous tasks. For example, driving while on the phone has been shown to result in slower responses to traffic signals (Strayer & Johnston, 2001). Other studies show that unfulfilled goals (like finishing a paper) interfere with tasks that require executive function (Marien et al., 2012; Masicampo & Baumeister, 2011b). Executive function includes working memory, reasoning, and problem solving, and can only pursue one goal at a time. Since executive function is in high demand during knowledge work, interleaving several long-term tasks (as studied in this paper) might decrease overall productivity when mental processes remain focused on prior goals. However, by consciously formulating plans for unfulfilled goals, such problems may be avoided (Masicampo & Baumeister, 2011a).

2.2. Task Switching

Task switching studies within cognitive sciences explore switching costs between tasks at a microscopic level (Kiesel et al., 2010), like the action of pressing a key. Such cognitive tasks require an appropriate configuration of mental resources, which task switching studies refer to as a 'task-set'. This is not to be confused with the task sets participants have worked on in the studies reported on in this paper, which cover higher-level real-world knowledge work. In a typical task switching experiment, effects of switching between two different task-sets are observed, like classifying a digit as odd or even, and classifying a letter as consonant or vowel. Switching costs include longer responses on the switched-to task immediately following the task switch. It is tempting to equate such a task switch with higher-level knowledge work where complex task goals need to be substituted. However, this is a fundamentally different operation (Monk et al., 2008).

2.3. Interruptions

Interruption studies measure the effects of short-lived secondary tasks interrupting a primary task, on both the primary and secondary task. A primary task here refers to higher-level goal-oriented work, like solving a Sudoku. Interruptions are short-term tasks which require suspension of the primary task, like answering a short question posed by a colleague. Interruptions can lead to annoyance and anxiety (Bailey et al., 2001), and feelings of stress and frustration (Mark et al., 2008, 2012). As summarized by Monk et al. (2008), some studies show people perform post-interruption tasks more slowly, and that more errors are made compared to pre-interruption performance. Characteristics determining the disruptiveness of interruptions include task similarity to the primary task, interruption complexity, control over interruption onset, availability of primary task retrieval cues, and duration (Monk et al., 2008). Not all interruptions are alike or as disruptive, depending on when they occur. Interruptions at task boundaries cause less anxiety and induce less errors (Bailey & Konstan, 2006), but this might depend on interruption relevance to the primary task (Gould et al., 2013). Other findings suggest resuming a primary task slowly can reduce the amount of errors made (Brumby et al., 2013). Based on such insights, interruption management systems attempt to alleviate the disruptiveness of interruptions by predicting interruptibility of the user (Turner et al., 2015).

When a secondary task interrupts a primary task two relevant time intervals become important to study: time taken (or allowed) to disengage from the primary task before starting work on the secondary task, and time taken to resume the primary task after completion of the secondary task (Altmann & Trafton, 2004; Boehm-Davis & Remington, 2009; Trafton et al., 2003). Although the study reported on in this paper does not include the traditional notion of a secondary task, the *disengagement and resumption stage* are still useful concepts, which can also be recognized within sequential multitasking studies while switching between two tasks.

2.4. Sequential Multitasking

In contrast to concurrent multitasking, sequential multitasking denotes the interleaving of several primary tasks which are executed one at a time. Compared to interruption studies there are no secondary tasks, as all tasks are long-lived and of equal importance. This is representative of common everyday knowledge work (González & Mark, 2004). Interruptions leading to task switches can either be internal (self-initiated) or external. Although most interruption studies focus on external interruptions, internal interruptions are as common in knowledge work, and there are different reasons for users to decide to switch tasks (Hardy & Gillan, 2012; Jin & Dabbish, 2009). Based on flow theory, internal interruptions can be categorized as originating from either negative (e.g., frustration, exhaustion) or positive (e.g., exploration, reorganization) feelings associated with the task (Adler & Benbunan-Fich, 2013).

Studies investigating sequential multitasking are among other things interested in measuring the effects of task interleaving on productivity and accuracy. The control condition typically consists of performing tasks in sequence, as opposed to an experimental condition where the same tasks need to be, or are voluntarily interleaved. Results show an inverted U-relationship between multitasking and productivity; there is thus an optimal amount of task switching which leads to the highest productivity. However, increased levels of multitasking lead to a significant loss in accuracy, indicating a trade-off between productivity and accuracy (Adler & Benbunan-Fich, 2012). More recently, subjective task difficulty has also been found to be a determining factor: easy tasks benefit from multitasking by increasing stimulation, whereas hard tasks decrease performance as the result of an overload in mental workload (Adler & Benbunan-Fich, 2015). Other studies show that users have a tendency to continue working on more rewarding tasks (with a continuous rate of return), and have a tendency to switch tasks after the completion of subgoals (Duggan et al., 2013; Payne et al., 2007). In addition, one of these studies shows that interruptions can disrupt task management (e.g., resuming an incorrect task) which can subsequently affect task performance (Duggan et al., 2013).

So far, however, no studies have investigated how different support for task switching in a desktop environment influences sequential multitasking. Prior studies generally employ a custom application where participants can switch between trivial tasks by the press of a button, e.g., solving a Sudoku, unscrambling letters to form words (Scrabble) and finding the "Odd One Out" between a set of shapes (Adler & Benbunan-Fich, 2012, 2015; Payne et al., 2007). Although these task sets make it straightforward to measure productivity and accuracy, they are not ecologically valid representations of real-world knowledge work within a desktop environment, in which a task often consists of multiple application windows and resources. In this paper we focus on how task switching as supported by dedicated workspaces influences task resumption and overall knowledge work, compared to a traditional desktop environment.

3. Method

Although the widespread use of virtual desktops clearly shows their importance, no prior studies have assessed their impact on knowledge work compared to a traditional desktop environment. The goal of this study is to *quantify the difference in time taken to switch between tasks under both environments, as well as measure possible effects on the performed tasks and experienced task load.* Similar to other sequential multitasking studies, participants work on similar tasks under two different conditions (Adler & Benbunan-Fich, 2012, 2015). Since our focus lies on the nature of task switches rather than degree of multitasking, we control for task switches by instructing participants to switch between given tasks at predetermined intervals (mandatory task interleaving), mimicking a heavy multitasking scenario with concurrent deadlines. Figure 1 shows an overview of a task switch within our experiment, indicating the resumption stage, which is the stage which is supported differently under both environments.

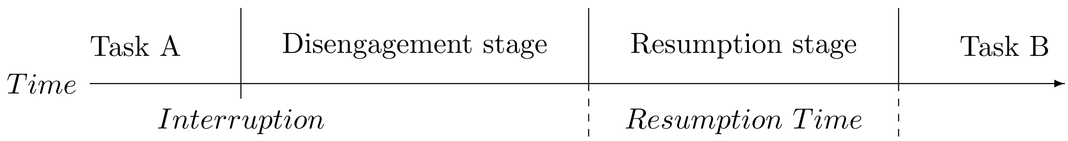

Figure 1: During the disengagement stage users wrap up a task after having received an interruption. During the resumption stage, users retrieve and prepare the next task to continue working on it. The resumption stage is supported differently under a traditional Windows 7 environment, than when using dedicated workspaces.

Adopting virtual desktops as dedicated workspaces allows instantaneous switching between parallel ongoing tasks, which in a traditional environment requires several operations. Therefore, we hypothesize *it takes longer to resume a previously suspended task under a traditional Windows 7 environment than when using dedicated workspaces.*

The necessary work to switch to the next task could be considered the equivalent of a secondary task within interruption studies. Unlike interruption studies, however, participants do not return to the same primary task but to a previously suspended goal. Since interruptions can lead to annoyance, anxiety, and stress (Bailey et al., 2001; Mark et al., 2008, 2012), we hypothesize *dedicated workspaces will reduce annoyance and anxiety compared to a traditional Windows 7 environment*.

Errors during task management have been shown to affect task performance (Duggan et al., 2013). Furthermore, in line with interruption duration and complexity influencing disruptiveness (Monk et al., 2008), the nature of a task switch might determine its disruptiveness. Therefore, we hypothesize *simplifying task switching by means of dedicated workspaces will improve task performance*.

3.1. Participants

Sixteen users (age 27–58, 12 male, 4 female) were recruited to participate in this study. Their backgrounds included program manager, ICT manager, pharmacist, clerk, student in social work, and software developer, representing a broad spectrum of knowledge workers. Only six participants worked within IT, ensuring not only expert computer users were recruited. All but one of the participants had extensive (> 1 year) experience with the Windows operating system, of which 12 specifically with Windows 7. The single inexperienced participant used OS X, and was not as familiar with the Windows 7 window manager, but did work within IT. Only four participants had used virtual desktops before. To incentivize participants in performing the given tasks, they were informed that the person ranking first in the study would win a cinema ticket. It was made clear that overall task score was calculated based on task progress, as well as accuracy. No other compensation was given for participating in the study.

3.2. Experimental Design

The experiment was run as a within-subjects design with one independent variable—*User Interface: Dedicated Workspaces Vs. Windows*. Under each of the two conditions participants worked on four separate tasks between which had to be switched at predetermined intervals. Two separate but similar task sets were created, one for each condition. We measured average *Resumption Time* while switching between tasks, as well as average task *Construction Time* needed to set up a task the first time. Other measured dependent variables were overall *Cognitive Load* throughout the experiment, and *Productivity* and *Accuracy* for each of the four tasks. The order in which the conditions were completed, and the order of the task sets used, were fully counterbalanced across participants (thus four different groups with four participants each, as shown in Table 1). In addition to counterbalancing for learning effects, this also accounts for differences in between the task sets.

3.3. Materials

We preferred for users to work in their own work environment to minimize the impact on our measures due to unfamiliarity with the workspace. However, due to practical limitations (e.g., work hours), six of the participants performed the study at a different desk, on

	Task A - B	Task B - A
Windows - DW	4	4
DW - Windows	4	4

Table 1: Counterbalancing of condition and task set order: Windows or Dedicated Workspaces (DW) first, and task set A or B first.

a notebook with a 15.6 inch screen provided by the experimenter. Although this introduced variability between work environments, in all cases the same setup was used under both *User Interface* conditions. No multiple monitors were used, but screen sizes did vary. The computer ran either an unmodified Windows 7 environment or was augmented using Laevo, a task-centric desktop interface offering dedicated workspaces in the form of virtual desktops (Jeuris et al., 2014). Within the context of this study, and in line with many virtual desktop implementations, a full screen overview simply displays different 'buttons' each representing an individual dedicated workspace (as shown in Figure 2) which is opened when clicked. Workspace representations (including a name, color, and icon) could be modified by the participants to help in identifying them.

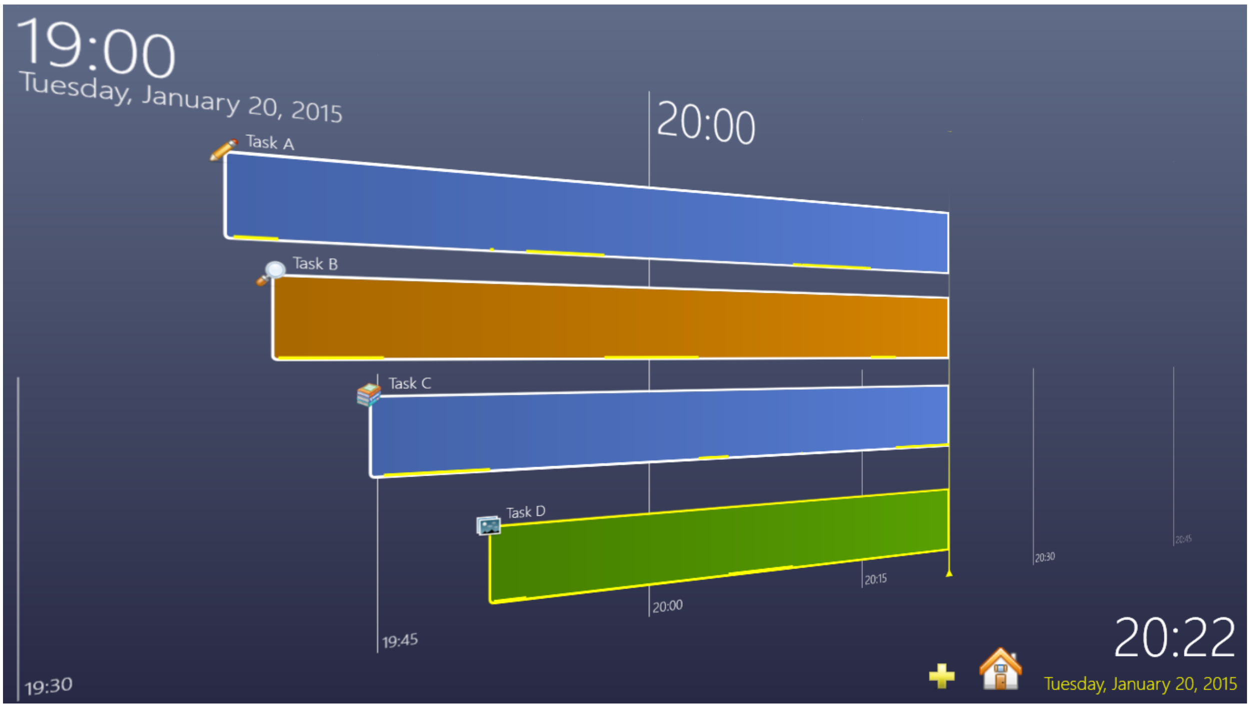

Figure 2: Example of the full screen overview used for task switching when using Laevo (near the end of the experiment), visualizing the different dedicated workspaces, represented as rectangular 'buttons' on a time line, which can be positioned vertically, and be given a name, color, and icon.

Participants had to switch between four tasks (listed below), which were designed according to criteria similar to earlier studies (Payne et al., 2007): performance should be measurable, and tasks should be linear in progression (should increase continuously and monotonically with time spent on the task). The tasks under both conditions were identical in nature, but contained different data. All tasks were designed to be long enough so they could not be completed. In contrast to earlier studies (Adler & Benbunan-Fich, 2012; Payne et al., 2007), we tried to ensure each unit of work over time (subtask) required the same amount of effort. Participants were instructed to work on subtasks in the order listed in the assignment. This allows for the measuring of productivity as a percentage of completed subtasks, and accuracy as a percentage of correctly performed work. Since the total amount of time spent on tasks is predetermined and is the same under both conditions, any difference in productivity or accuracy can be accounted for by the condition under which the task was performed. These task requirements eliminate the possibility of using more complex tasks like solving a Sudoku. However, as opposed to prior sequential multitasking studies, we did not simulate or modify any of the applications used throughout the experiment, but rather let the participants work in an unmodified desktop environment, as shown in Figure 3. This supports our main intent: *simulating a representative desktop environment during heavy multitasking, where several application windows and resources are open simultaneously.*

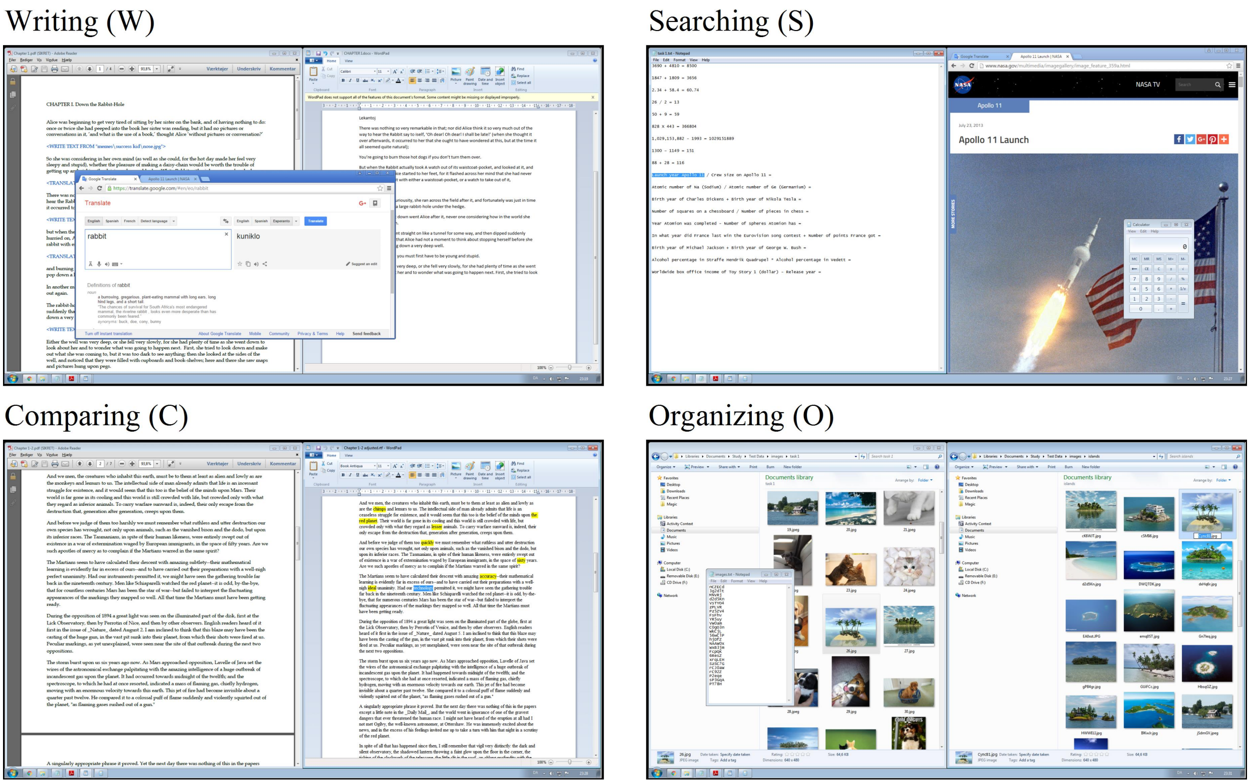

Figure 3: For each of the four tasks, a screenshot is depicted showing one possible window configuration while working on them using a separate dedicated workspace for each. Note that only windows required for the task at hand are represented on the Windows task bar, due to the use of virtual desktops.

Writing (W): Participants type text[1] found within a locked PDF file (copy/paste is dis-

[1]Alice in Wonderland by Lewis Carroll, Chapter 1 and 3

abled) into a new text document. No formatting needs to be applied. The text includes assignments at regular intervals asking participants to substitute the assignment with: text displayed in an image located on the hard drive, or with the translation of a word using Google Translate. This implies four applications need to be used: a file explorer window, a browser, PDF reader, and a text processor (see Figure 3, W). Task productivity is measured using the number of written characters and accuracy using the number of correctly written characters.

Searching (S): Participants complete calculations by searching the Internet and adding both the intermediate and final results to the assignment text file. For example, "*Height of the Eifel Tower in meters + year when it was completed =*". This implies three applications need to be used: a browser, a text processor, and a calculator (see Figure 3, S). Task productivity and accuracy are based on three points per calculation; one point for each retrieved number and a third point for the correctness of the calculation itself. In case there are different sources stating different numbers, any of them was considered to be correct. Points for the calculation were still given when using wrong numbers but performing a correct calculation.

Comparing (C): Participants highlight differences between an original and modified text document[2]. Modifications include synonyms, left out words, or additional words, but the modified text is still grammatically correct. This implies two applications need to be used, specifically two text processors (see Figure 3, C). Task productivity is measured by the last identified difference, and accuracy is measured by any skipped or incorrectly highlighted difference over the total amount of differences identified.

Organizing (O): To mimic folder navigation, a folder hierarchy contains images organized by type of object (e.g., bridges, islands ...). A task folder contains the same images but disorganized. Participants need to identify what is displayed in the disorganized image, find the corresponding folder, within it find the image, and subsequently copy its filename into a text document. This implies two applications need to be used: a file explorer window and a text processor (see Figure 3, O). However, experienced users generally use several file explorer windows. In addition, the image viewer might be used when image thumbnails are unclear. Task productivity is measured by the number of images retrieved and accuracy by the number of skipped images, or erroneous entries in the text document.

In addition, for each task a text file with a short assignment description was provided, referring to the required file and folder location needed for that task. Using the traditional desktop environment while running the experiment, this results in a workspace as shown in Figure 4, representative of heavy multitasking. In contrast, Figure 3 shows what individual tasks look like when using dedicated workspaces, whereas Figure 2 shows the interface used to switch between them. The main difference lies in how many windows are accessible from the Windows task bar.

[2]The War of The Worlds by H. G. Wells, Chapter 1–5

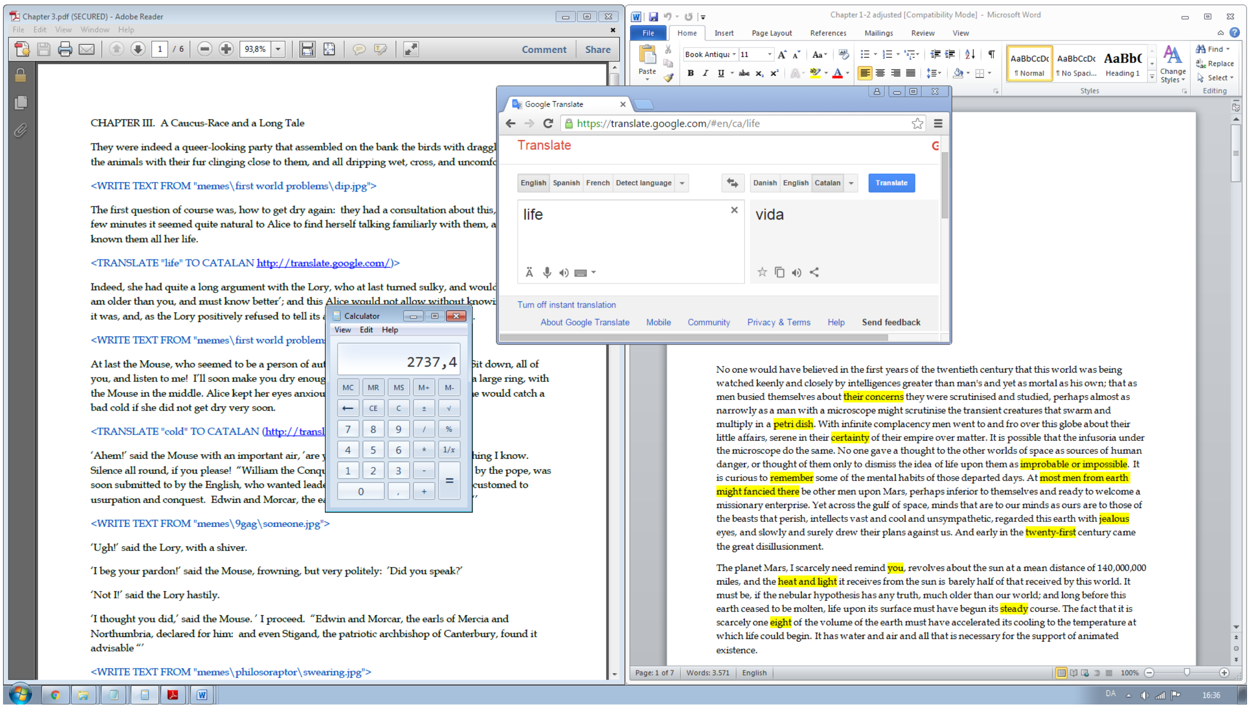

Figure 4: Screenshot showing a possible window configuration while working on the tasks under a traditional desktop environment. Note that windows required for all four tasks are represented on the Windows task bar.

3.4. Procedure

Six users had installed and used Laevo for at least a full day prior to the experiment, and thus did not need an introduction to the system. The other 10 users were given a one-hour introduction on how to use Laevo, focusing primarily on hands-on experience with the virtual desktop functionality. Given that this is the only functionality of Laevo which was used as part of the experiment, this provided them with sufficient training to take part in the study. Before starting the experiment, participants received a 10 minute introduction of the setup and tasks, and were asked about their background, experience with Windows 7, and whether they had used virtual desktops before. Tasks were demonstrated using an example task set, after which participants were asked to work on them shortly. In both conditions, participants worked on the tasks for 50 minutes, after which they completed the NASA-TLX test to assess overall cognitive load during the experiment. A modified version of the NASA-TLX test was used, Raw TLX (Hart, 2006), which eliminates the weighting process of the separate subscales. In between both conditions, 10 minutes were reserved to recuperate from the heavy workload. In total, the duration of the experiment was two hours for participants that had used Laevo before, and three hours for those that still needed an introduction. During the study, the experimenter notified the user when to switch between tasks by stating the task number, and a short description. E.g., "*Now please switch to task A, which is the writing task.*" These notifications were given at predetermined intervals of 2, 4.5 and 6 minutes, totaling in 12.5 minutes of work per task, representative of real-world knowledge work (González & Mark, 2004). The sequence of tasks W, S, C and O is depicted in Figure 5.

W	S	C	O	W	S	C	O	W	S	C	O	$Task$
2	4.5	6	2	4.5	6	2	4.5	6	2	4.5	6	$Min.$

Figure 5: Task sequence under both conditions.

A stopwatch ran throughout the experiment, used to keep track of task switches. In rare occasions when a problem occurred, either with the dedicated workspaces environment, or participants raising questions about the task set, the timer was paused until the issue was resolved. When instructed to switch to a new task, participants were allowed to wrap up the task they were working on; they were allowed to finish subtasks, e.g., a copy/paste operation, or finish writing a sentence, but were not allowed to commence work on a new subtask. Any exceptionally late responses were addressed by restating that the participant had to switch to the next task. Finishing subtasks was allowed to control for disruptive effects due to differences within the *disengagement stage*, since interruptions at task boundaries are know to be less disruptive (Bailey & Konstan, 2006). Within this study, we are only interested in effects due to differences within the *resumption stage* as shown in Figure 1 (Boehm-Davis & Remington, 2009). Both time taken during the disengagement and resumption stage were measured by the observing experimenter for all but three of the participants (due to three experiments which ran with two participants in parallel) by marking the time when the participant stopped working on the previous task, and the time when the participant performed the first operation on the switched-to task. Once all windows required to continue work on the task were retrieved and readily accessible, any mouse or keyboard input on a window of the switched-to task was considered to be the first operation. The resulting time intervals correspond to measured time intervals within interruption studies (Altmann & Trafton, 2004; Monk et al., 2008; Trafton et al., 2003). The first time participants were asked to work on a task (and are thus not resuming it), we measured construction time in the same way: the time taken to open the required documents and to set up the workspace.

4. Results

4.1. Task Resumption and Construction Times

Figure 6 shows an overview of average resumption times and standard deviation per task switch throughout the experiment for both conditions. We used a paired one-tailed t-test to analyze overall average *Resumption Time* in both conditions. We found a significant difference ($t(12) = -2.95, p < 0.01$) between *Dedicated Workspaces* ($\mu = 7.52, SD = 2.93$) and *Windows* ($\mu = 26.78, SD = 25.61$). The effect size using Glass's $\Delta = 0.75$.

Since we noticed a high variance under the *Windows* condition (traditional desktop interface), we investigated the individual resumption times of participants in more detail. Figure 7 shows an overview of average resumption times and standard deviation per participant for both conditions, ordered by average resumption time under the *Windows* condition.

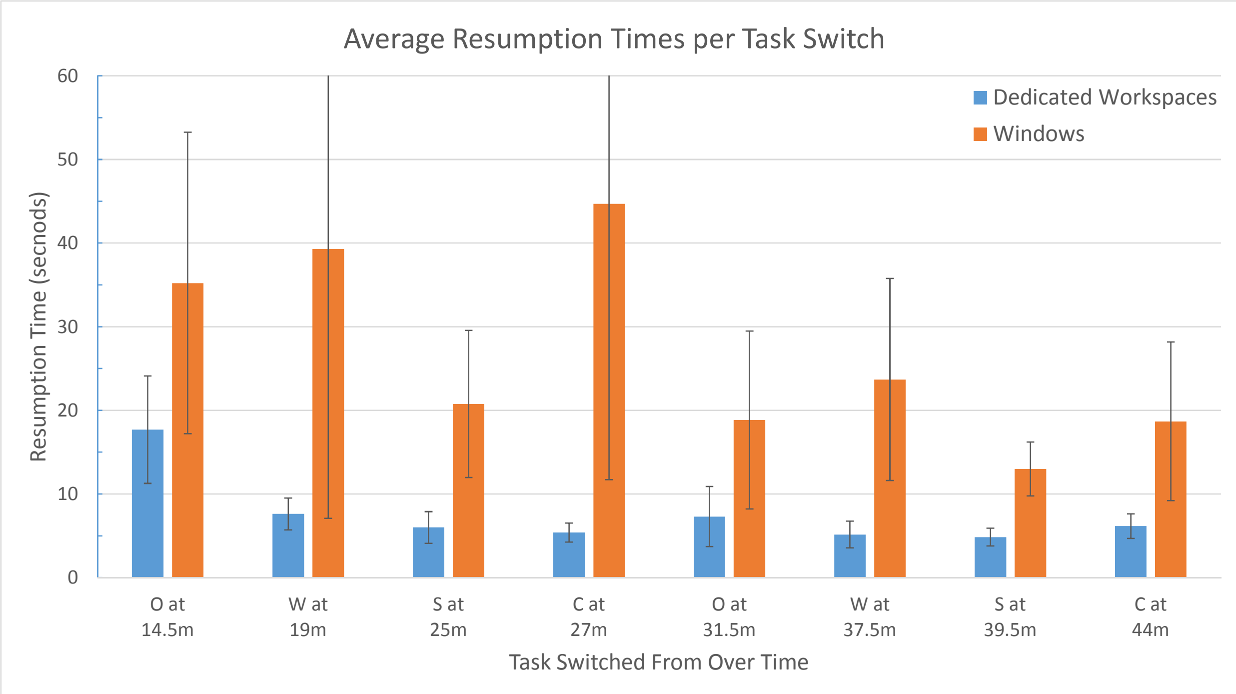

Figure 6: Overview of average resumption time per individual task switch under both conditions.

Three outliers can be recognized with resumption times under the *Windows* condition which vary considerably more than the other participants; all three outliers (which includes the OS X user) encountered a task switch which took them over two minutes to complete (250, 140, and 203 seconds respectively). Removing these outliers and rerunning the t-test we measure a more significant difference ($t(9) = -4.43, p < 0.001$) between *Dedicated Workspaces* ($\mu = 6.26, SD = 1.92$) and *Windows* ($\mu = 15.79, SD = 7.00$). The effect size using Glass's $\Delta = 1.36$. On average *Resumption Time* is 9.53 seconds shorter when using *Dedicated Workspaces*.

We used a paired one-tailed t-test to analyze average *Construction Time* in both conditions, expecting additional time needed to set up *Dedicated Workspaces*. We found a significant difference ($t(12) = 3.15, p < 0.005$) between *Dedicated Workspaces* ($\mu = 27.21, SD = 17.05$) and *Windows* ($\mu = 14.05, SD = 5.43$). The effect size using Glass's $\Delta = 2.43$. On average *Construction Time* is 13.16 seconds longer when using *Dedicated Workspaces*.

Figure 8 shows an overview of all task switches throughout the experiment for participants one through ten, corresponding to the participants listed in Figure 7. The first three task switches represent construction times where the workspace still needs to be set up. The remaining task switches represent resumption times. The y-axis shows total (cumulative) time spent on task switches over time when using dedicated workspaces, minus time spent on the same task switches under the traditional Windows environment. The graphs thus show whether, and when, using dedicated workspaces resulted in spending more, or less time on task switching. The increase in *Construction Time* when using *Dedicated Workspaces* can

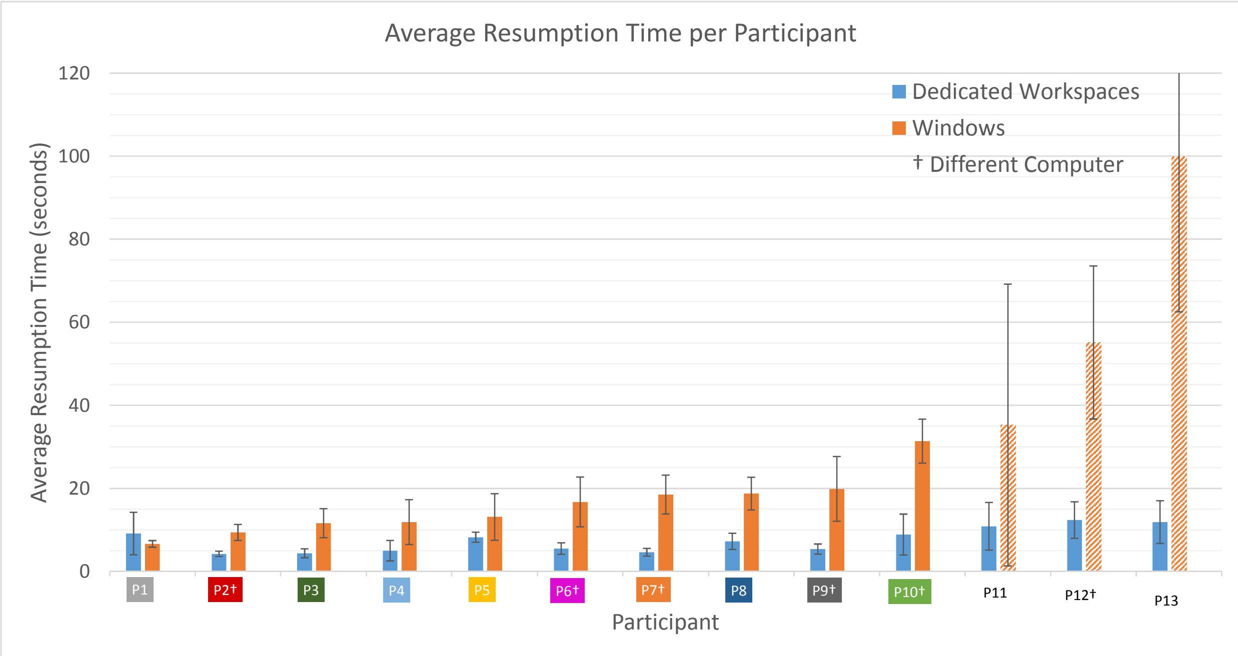

Figure 7: Overview of average resumption times per participant under both conditions. P11, P12, and P13 were identified as outliers due to proportionally high variance in the Windows condition compared to the other participants. Participants marked with '†' did not perform the study in their own work environment, and did not use their own computer.

be seen during the three initial task switches. However, during subsequent task switches, *Resumption Time* is considerably lower. Therefore, for all but two participants, *Dedicated Workspaces* resulted in spending less time overall on task switching at the end of the experiment. Seven out of ten participants already made up for additional construction time when using *Dedicated Workspaces* after just three subsequent task resumptions.

4.2. Cognitive Load

Cognitive load was measured using Raw TLX (Hart, 2006), which assesses work load on five separate scales with 21 gradations. The overall measured task load index is the average of the five scales. An overview of averages and standard deviations of the separate TLX scales is shown in Figure 9. We used a paired one-tailed t-test to analyze overall *Cognitive Load* in both conditions, expecting a reduced cognitive load when using *Dedicated Workspaces*. We found a significant difference ($t(15) = -2.45, p < 0.05$) between *Dedicated Workspaces* ($\mu = 9.66, SD = 2.48$) and *Windows* ($\mu = 11.25, SD = 2.91$). The effect size using Hedges's $g_{av} = 0.56$. On average *Cognitive Load* is rated 14.1% lower when using dedicated workspaces.

Running two-tailed t-tests on the separate scales results in only finding a significant effect ($p < 0.05$) for *Temporal Demand* ($t(15) = -2.41, p < 0.05$) and *Effort* ($t(15) = -3.30, p < 0.01$). For *Temporal Demand* the mean of *Dedicated Workspaces* ($\mu = 11.25, SD = 4.19$) was lower than *Windows* ($\mu = 13.88, SD = 3.30$), Hedges's $g_{av} = 0.66$. On average *Temporal*

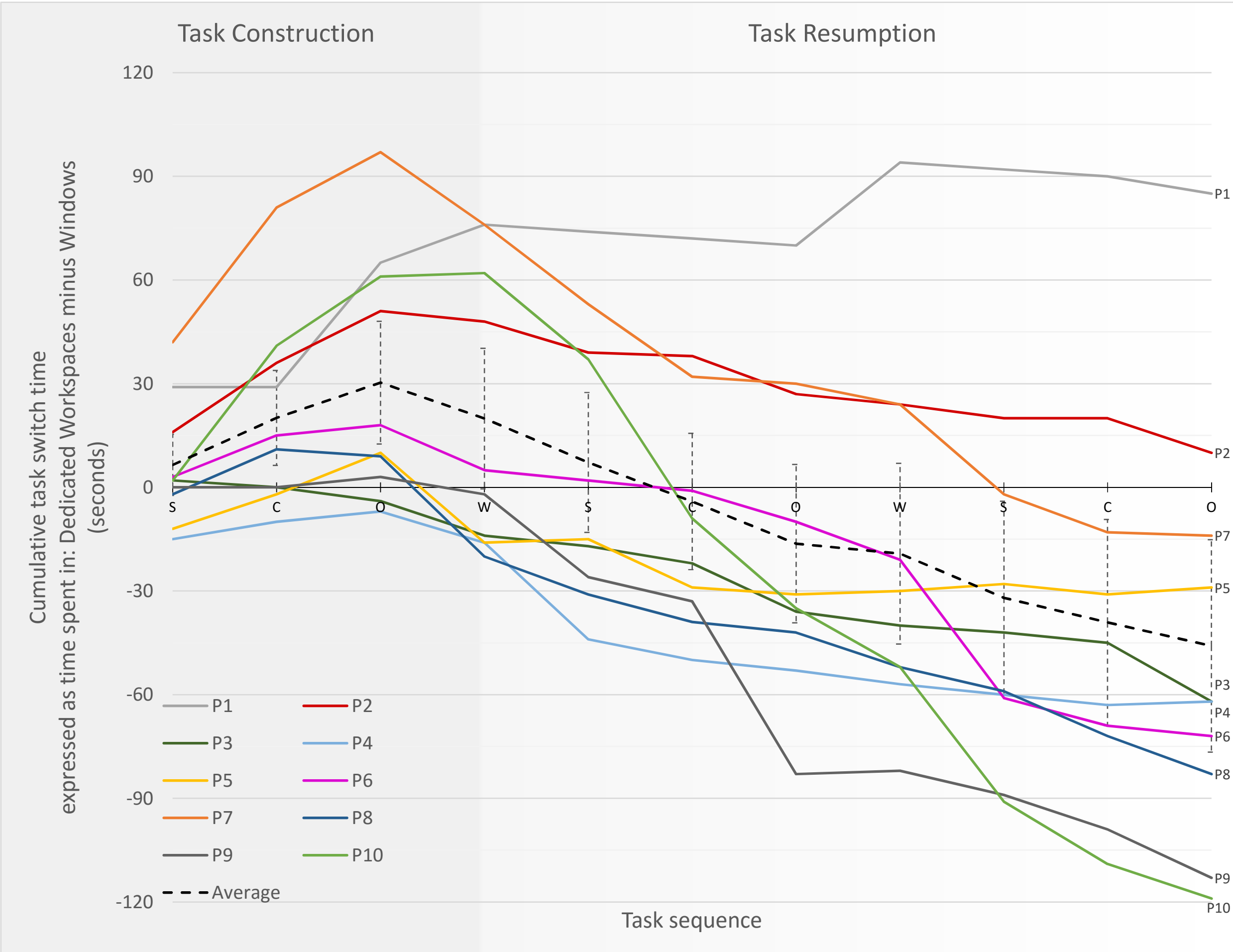

Figure 8: An overview of task switches throughout the experiment per participant. The first three task switches represent construction times where the workspace still needs to be set up. The remaining task switches represent resumption times. The y-axis shows total (cumulative) time spent on task switches over time when using dedicated workspaces, minus time spent on the same task switches under the traditional Windows environment.

Demand is rated 23.4% lower when using dedicated workspaces. For *Effort* the mean of *Dedicated Workspaces* ($\mu = 11.75, SD = 3.44$) was lower than *Windows* ($\mu = 13.10, SD = 3.34$), Hedges's $g_{av} = 0.38$. On average *Effort* is rated 11.5% lower when using dedicated workspaces.

4.3. Task Productivity and Accuracy

We used paired two-tailed t-tests to analyze task *Productivity* and *Accuracy* in both conditions for all four tasks ($N = 16$). There are no significant effects ($p < 0.05$) between *Dedicated Workspaces* and *Windows* for any of the tasks on *Productivity* or *Accuracy*, of which an overview is provided in Figure 10.

To inspect the severity of learning effects we used paired one-tailed t-tests to analyze task *Productivity* and *Accuracy* of tasks between *Session 1* and *Session 2*: the first, and

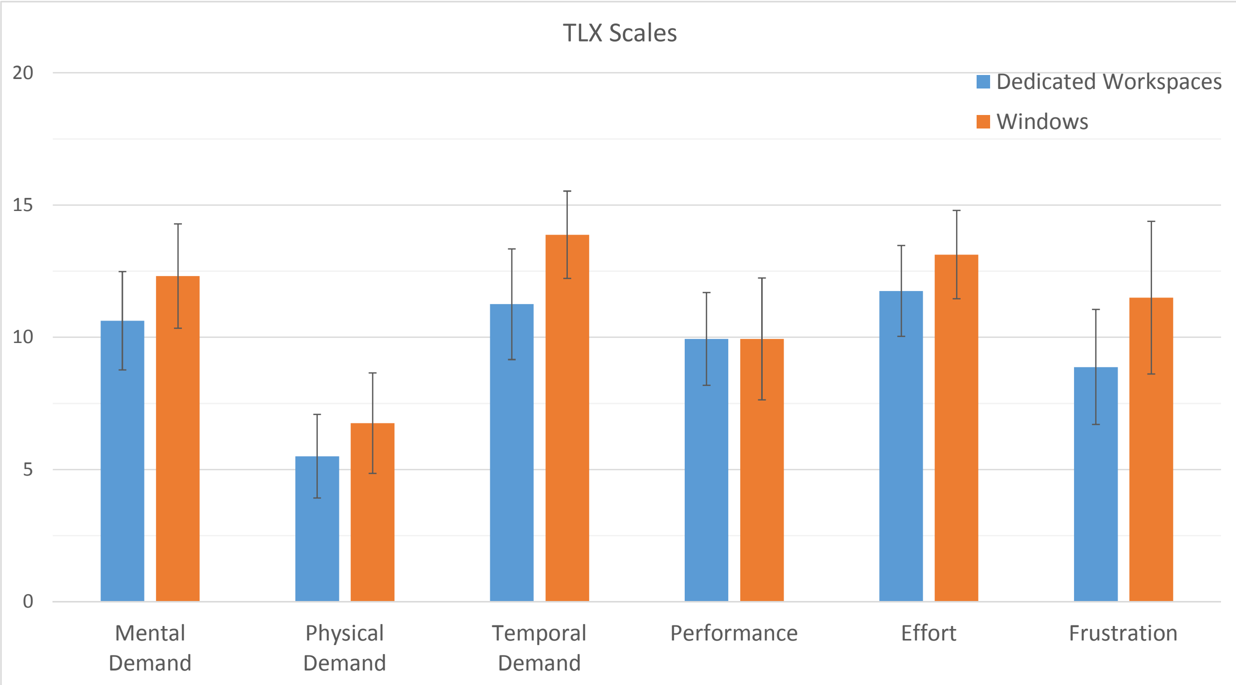

Figure 9: Breakdown of averages for each scale used as part of the Raw TLX test.

	Writing		Searching		Comparing		Organizing	
	Prod.	Acc.	Prod.	Acc.	Prod.	Acc.	Prod.	Acc.
p	0.51	0.63	0.63	0.24	0.66	0.80	0.39	0.29
t	0.67	0.49	0.50	1.21	0.44	0.26	0.88	1.09
DW µ	0.48	0.99	0.42	0.83	0.45	0.84	0.46	0.92
Win µ	0.49	0.99	0.40	0.79	0.43	0.83	0.43	0.89
DW SD	0.21	0.01	0.20	0.13	0.21	0.11	0.26	0.13
Win SD	0.23	0.01	0.20	0.09	0.17	0.12	0.24	0.20

Figure 10: Overview of t-tests comparing productivity and accuracy of all tasks between dedicated workspaces (DW) and a traditional desktop environment (Win).

second condition participants were exposed to respectively. We found a significant difference between *Session 1* and *Session 2* for *Productivity* on all four tasks, as summarized in Figure 11. On average participants performed 20.8% more work overall in *Session 2* than in *Session 1*.

5. Discussion

5.1. Task Resumption and Construction Time

Despite the fact that dedicated workspaces are specifically designed to facilitate task switching, no prior studies have quantified task resumption time with or without dedicated workspaces. Our study shows that it takes on average 10 seconds longer to switch between

	Writing		Searching		Comparing		Organizing	
	Prod.	Acc.	Prod.	Acc.	Prod.	Acc.	Prod.	Acc.
p	< 0.001	0.44	< 0.001	0.21	< 0.001	0.13	< 0.001	0.19
t	4.25	0.79	4.91	1.32	4.26	1.62	4.28	1.36
Hedges g_{av}	0.30	-	0.41	-	0.42	-	0.41	-
S1 μ	0.45	0.99	0.37	0.79	0.40	0.82	0.39	0.89
S2 μ	0.52	0.99	0.45	0.83	0.48	0.85	0.50	0.92
S1 SD	0.20	0.01	0.19	0.12	0.17	0.12	0.23	0.20
S2 SD	0.23	0.01	0.20	0.10	0.20	0.11	0.25	0.13

Figure 11: Overview of t-tests comparing productivity and accuracy of all task between the first session (S1) and the second session (S2) participants took part in.

tasks under a traditional Windows 7 environment, than when using dedicated workspaces. Our study thus provides first empirical evidence *to which degree* dedicated workspaces improve task resumption time. Based on these findings, we identify that not all users benefit from dedicated workspaces equally, as indicated by the big differences in task resumption times under both conditions (see Figure 7).

Participants P11, P12, and P13 encountered task switches lasting over two minutes using the traditional Windows 7 environment, indicating a complete breakdown in finding the required resources to resume a previously suspended task. One of these participants (P12) was an inexperienced Windows user (OS X user), unfamiliar with some of the features of Windows 7. However, the other two participants were a consultant and clerk, using Windows 7 on a regular basis as part of their work. The consultant (P13) at times expressed severe annoyance when being unable to retrieve the necessary resources, after which she would clean up her workspace entirely by closing all windows and start over. The clerk (P11) struggled particularly when navigating between folders (using Windows explorer), and sometimes had to reopen the assignment description to look up the required path for the task. In line with these observations, we anticipate that users often close resources because of the overhead of keeping them open; our study shows empirical evidence that severe detrimental effects can occur when not doing so. Dedicated workspaces alleviate this problem by hiding irrelevant resources that might interfere while switching between tasks.

In contrast, what seemingly sets more ‘expert’ users apart is they take great care in making the work environment their own. They introduce additional cues which allow them to disambiguate between tasks. This is in line with observations made by Kidd (1994): “*the marks are on the knowledge worker*”. For example, P1 gave more meaningful names to documents so they could easily be recognized on the task bar during task switching. Possibly such actions are indicative of formulating plans to return to the task at a later moment in time. This has been shown to alleviate the detrimental effects of unconsciously remaining focused on unfulfilled goals (Masicampo & Baumeister, 2011a). Rather than reusing a single Windows explorer window, some users opened multiple instances with different file paths open in them, positioned optimally for each individual task. They thus seem to be able to keep more resources open, yet easily keep track of which tasks they belong to. Future studies should investigate in more detail what task resumption and resumption time in a desktop environment is comprised of (with more granularity) in order to gain a better understanding of where the exact differences between users lie.

Dedicated workspaces allow the user to set up a dedicated environment for each individual goal-oriented task, which is associated with a longer set-up time (13 seconds on average) within the user interface we used (Jeuris et al., 2014). However, as demonstrated in Figure 2, an intrinsic quality of dedicated workspaces is task switching is almost completely reduced to identifying a task definition among a list of open parallel tasks. At the end of our experiment all but two participants spent less time on task switching when using dedicated workspaces than when using a traditional Windows 7 environment (Figure 8); for the remaining two participants the trend predicts a similar result given a longer experiment. As P2 stated: "*[dedicated workspaces] would mainly be useful for tasks I haven't worked on for a longer period of time. Since [during the experiment] there is not much time in between task switches, using Windows 7, it is not that much more difficult.*" Providing support for dedicated workspaces opens up the opportunity for rapid access to a multitude of long-lived workspaces (rather than just four), which we hypothesize would greatly benefit even experienced users.

5.2. Cognitive Load

Raw TLX (Hart, 2006) was used to measure cognitive load, meaning our significant result reflects our participants' subjective preference for dedicated workspaces over a traditional desktop environment during heavy multitasking. This is in line with our hypothesis of task switch interruptions being more disruptive (causing more annoyance and anxiety) under a traditional desktop environment. Interestingly, it are mainly *effort* (how hard did you have to work to accomplish your level of performance) and *temporal demand* (perceived time pressure—was the pace slow or rapid) which influenced cognitive load. The difference in reported effort can be explained by the additional required operations to switch between tasks under a traditional desktop environment. However, this does not explain why participants felt more rushed; under both conditions participants worked on the same tasks, and the incentive remained unchanged. Information overload might be at play here. As Mulder et al. (2006) stated: "*the heart of what information overload really is may very well lie between tasks rather than within*". Since information overload can cause stress, the mere visibility of previously suspended tasks could potentially increase perceived time pressure. Even participants fluent in task switching under a traditional desktop environment seem to confirm this, e.g., P4 stated having more "*breathing space*" when using dedicated workspaces. In contrast, participants were not significantly more frustrated under a traditional desktop environment, contradicting the finding of our overall measure. However, quite a few of the participants seemed to be accustomed to heavy multitasking, thus struggling minimally with task switches.

Given our experimental setup, we do not find that the NASA-TLX test provides sufficient sensitivity in order to easily interpret effect size. Future work could investigate how to obtain and interpret objective measurements of cognitive load while performing knowledge work in a desktop environment. There is an opportunity to use biometric data; skin conductance, blood volume, and pulse rate in particular seem promising (Janssen et al., 2015; Mark et al., 2012), which are indicative of *arousal*. Arousal however is not indicative of *valence*—the pleasantness or unpleasantness of stimuli. Therefore we first need a better understanding of the different triggers which influence arousal during knowledge work, in order to assess whether they are indicative of cognitive load or not.

5.3. Task Productivity and Accuracy

Interruption studies indicate that more errors are made right after an interruption, and that longer, more demanding interruptions lead to longer resumption times (Monk et al., 2008). Considering that the required configuration work to switch between tasks could be interpreted as interruptions during knowledge work, albeit between two distinct tasks, we hypothesized a measurable effect on task performance. However, due to the systematic nature of the task sets used, severe learning effects were observed (see Figure 11). The resulting high variance in task productivity and accuracy eliminated the opportunity to measure any significant effects. Rather than trying to make task sets measurable, future studies could intersperse them with traditional cognitive tests (e.g., Stroop test, n-back) on which to measure progress and accuracy instead. This would allow participants to work on even more representative task sets. On the other hand, this would introduce additional interruptions which need to be controlled for under both conditions. An alternative approach could be, rather than measuring overall task performance, only measuring task performance during the interval immediately following task switches. Although differences in cognitive load can impact overall performance, interruption studies predict a bigger impact immediately following the interruption (Monk et al., 2008).

5.4. Threats to Validity

Our choice of letting participants work within a real desktop environment (in order to increase ecological validity with respect to the resumption stage) imposed us with several practical limitations. Task resumption time measures could not be automated, but needed to be measured by a researcher using a stopwatch. This implies inaccuracies due to the experimenter's response time and possible bias. To prevent bias, a protocol was defined up front which defines what constitutes task resumption, based on prior research (Altmann & Trafton, 2004; Monk et al., 2008; Trafton et al., 2003). To optimize accuracy, rather than starting and stopping the stopwatch to measure time intervals, only instants in time were noted. This allowed the experimenter to easily update a measure when a later observation indicated a task was not yet resumed. Time intervals were only calculated after a task switch had occurred. In contrast to automated measures, a benefit is intent can be interpreted as well, as not all mouse or keyboard input on a task set implies task resumption; one application window, part of the task set, could be correctly retrieved and receive input, but a secondary window might still be needed which takes the participant some additional time to retrieve. In addition, verbal utterances by participants often provided rich cues about the exact time a task was resumed. Three experiments were conducted monitoring two participants simultaneously. During these experiments the researcher could only focus on measuring task resumption time for one of the participants, resulting in 13 (as opposed to 16) task resumption time measures.

One of the core challenges in setting up a laboratory experiment assessing productivity and accuracy of knowledge work, as influenced by a manipulated variable, is how to control for the knowledge work itself. Allowing participants to work freely on their actual work would provide maximum validity. This, however, makes any comparison across tasks impossible due to the 'non-routine' problem solving nature of knowledge work (Kidd, 1994). In such a setup, productivity and accuracy can vary widely from one session to another. Creating comparable task sets is a trade-off between the sensitivity of task performance measures, and the ecologically valid circumstances we strive for as part of our methodology (González & Mark, 2004). More concretely, equally spaced short subtasks simplify measuring performance and allow for shorter experiments, but are not as representative of knowledge work. In contrast, longer subtasks can be more ecologically valid, but by their nature will introduce more variance in performance measures. Using our task sets, we experienced that the effects would have to be of a considerable size in order to be measurable when running a study with a limited number of participants.

6. Conclusion

Although prior research adopted dedicated workspaces with the premise that they simplify multitasking, there is no experimental evidence of their impact. In this paper, we measured the task resumption time, construction time, self-reported cognitive load, task productivity, and accuracy while multitasking under both a traditional desktop environment and one augmented by dedicated workspaces. Our study showed that the time to switch between tasks, and cognitive load, is considerably higher in a traditional desktop environment than when using dedicated workspaces. It takes on average 10 seconds longer to switch between tasks under a traditional desktop environment. The time lost on setting up a workspace is thus almost instantly compensated for by faster task resumption. However, not all users experience the same decrease in task resumption time. More experienced users require several revisitations to make up for the time lost during task construction.

These results are in line with results from related interruption studies, observational studies of desktop environments, and theories within cognitive science, and thus provide a strong argument for supporting dedicated workspaces within desktop environments. In addition, related studies predict using dedicated workspaces could increase task accuracy and productivity, but due to a trade-off between ecological validity and task sensitivity our study was unable to measure a significant effect. Further studies are needed to evaluate the impact on task performance.

References

Adler, R. F., & Benbunan-Fich, R. (2012). Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies*, *70*, 156–168.

Adler, R. F., & Benbunan-Fich, R. (2013). Self-interruptions in discretionary multitasking. *Computers in Human Behavior*, *29*, 1441–1449.

Adler, R. F., & Benbunan-Fich, R. (2015). The effects of task difficulty and multitasking on performance. *Interacting with Computers*, *27*, 430–439.

Altmann, E. M., & Trafton, J. G. (2004). Task interruption: Resumption lag and the role of cues. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* CogSci '04 (pp. 43–48). Lawrence Erlbaum Associates, Inc.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, *22*, 685–708.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001). The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Proceedings of IFIP TC.13 International Conference on Human-Computer Interaction* INTERACT '01 (pp. 593–601). IOS Press.

Bardram, J., Bunde-Pedersen, J., & Soegaard, M. (2006). Support for activity-based computing in a personal computing operating system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '06 (pp. 211–220). ACM.

Beale, R., & Edmondson, W. (2007). Multiple carets, multiple screens and multi-tasking: New behaviours with multiple computers. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 1* BCS-HCI '07 (pp. 55–64). British Computer Society.

Boehm-Davis, D. A., & Remington, R. (2009). Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention*, *41*, 1124–1129.

Brumby, D. P., Cox, A. L., Back, J., & Gould, S. J. (2013). Recovering from an interruption: Investigating speed- accuracy trade-offs in task resumption behavior. *Journal of Experimental Psychology: Applied*, *19*, 95–107.

Duggan, G. B., Johnson, H., & Sørli, P. (2013). Interleaving tasks to improve performance: Users maximise the marginal rate of return. *International Journal of Human-Computer Studies*, *71*, 533–550.

González, V. M., & Mark, G. (2004). "Constant, constant, multi-tasking craziness": Managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '04 (pp. 113–120). ACM.

Gould, S. J., Brumby, D. P., & Cox, A. L. (2013). What does it mean for an interruption to be relevant? an investigation of relevance as a memory effect. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 149–153). SAGE Publications volume 57.

Grudin, J. (2001). Partitioning digital worlds: Focal and peripheral awareness in multiple monitor use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '01 (pp. 458–465). ACM.

Hardy, M., & Gillan, D. J. (2012). Voluntary task switching patterns in everyday tasks of different motivational levels. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 2128–2132). SAGE Publications volume 56.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* HFES '06 (pp. 904–908). SAGE.

Henderson, D. A., Jr., & Card, S. (1986). Rooms: The use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics*, *5*, 211–243.

Houben, S., Bardram, J. E., Vermeulen, J., Luyten, K., & Coninx, K. (2013). Activity-centric support for ad hoc knowledge work: A case study of co-activity manager. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '13 (pp. 2263–2272). ACM.

Hutchings, D. R., Smith, G., Meyers, B., Czerwinski, M., & Robertson, G. (2004). Display space usage and window management operation comparisons between single monitor and multiple monitor users. In *Proceedings of the Working Conference on Advanced Visual Interfaces* AVI '04 (pp. 32–39). ACM.

Janssen, C. P., Gould, S. J., Li, S. Y., Brumby, D. P., & Cox, A. L. (2015). Integrating knowledge of multitasking and interruptions across different perspectives and research methods. *International Journal of Human-Computer Studies*, *79*, 1–5.

Jeuris, S., Houben, S., & Bardram, J. (2014). Laevo: A temporal desktop interface for integrated knowledge work. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*

UIST '14 (pp. 679–688). ACM.
Jin, J., & Dabbish, L. A. (2009). Self-interruption on the computer: A typology of discretionary task interleaving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '09 (pp. 1799–1808). ACM.
Kidd, A. (1994). The marks are on the knowledge worker. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '94 (pp. 186–191). ACM.
Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switchinga review. *Psychological bulletin*, *136*, 849–874.
Marien, H., Custers, R., Hassin, R. R., & Aarts, H. (2012). Unconscious goal activation and the hijacking of the executive function. *Journal of personality and social psychology*, *103*, 399–415.
Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: More speed and stress. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '08 (pp. 107–110). ACM.
Mark, G., Voida, S., & Cardello, A. (2012). "A pace not dictated by electrons": An empirical study of work without email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '12 (pp. 555–564). ACM.
Masicampo, E., & Baumeister, R. F. (2011a). Consider it done! plan making can eliminate the cognitive effects of unfulfilled goals. *Journal of personality and social psychology*, *101*, 667–683.
Masicampo, E., & Baumeister, R. F. (2011b). Unfulfilled goals interfere with tasks that require executive functions. *Journal of Experimental Social Psychology*, *47*, 300–311.
Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied*, *14*, 299–213.
Mulder, I., de Poot, H., Verwij, C., Janssen, R., & Bijlsma, M. (2006). An information overload study: Using design methods for understanding. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* OZCHI '06 (pp. 245–252). ACM.
Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General*, *136*, 370–388.
Ringel, M. (2003). When one isn't enough: An analysis of virtual desktop usage strategies and their implications for design. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* CHI EA '03 (pp. 762–763). ACM.
Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '09 (pp. 1819–1828). ACM.
Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, *12*, 462–466.
Trafton, J., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, *58*, 583–603.
Turner, L. D., Allen, S. M., & Whitaker, R. M. (2015). Interruptibility prediction for ubiquitous systems: Conventions and new directions from a growing field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* UbiComp '15 (pp. 801–812). ACM.
Voida, S., Mynatt, E. D., & Edwards, W. K. (2008). Re-framing the desktop interface around the activities of knowledge work. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* UIST '08 (pp. 211–220). ACM.